



## Lectura 5: Pruebas no paramétricas con proporciones

En este capítulo se hablará de algunas pruebas que para inferir acerca de hipótesis nula o alternativas, no mencionan parámetro alguno, no hacen suposiciones ni nada, a estas pruebas se les llama no paramétrica o libres de distribución

### Prueba Exacta de Fisher

Permite analizar la asociación entre dos variables dicotómicas basándose en los conteos observados y registrados en una matriz de  $2 \times 2$ , para variables binarias (sí o no) (1 o 0).

Por ejemplo dada la siguiente tabla

		Variable 1		Total
		Presente	Ausente	
Variable 2	Presente	a	b	a+b
	Ausente	c	d	c+d
Total		a+c	b+d	n

luego, si se asume independencia entre ambas variables, la probabilidad exacta de observar el conjunto de frecuencias de la tabla, corresponde a una distribución hipergeométrica

$$P(\text{Tabla}) = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!}$$

De esta forma si encontramos todas las posibles tablas de contingencia, entonces podemos (sumando), determinar la probabilidad exacta, sumando las probabilidades que sean **menor o igual** a la probabilidad calculada en la primera tabla, tenemos que hacer una hipótesis nula y alternativa las cuales generalmente siempre tomarán la forma de

- $H_0$ : las variables son independientes (no existe asociación)
- $H_a$ : las variables están relacionadas (existe asociación)

En R, `fisher.test(x, y=NULL, alternative="two.sided", conf.level=0.95)`, en donde **x** tiene que ser la tabla de contingencia, si no, **x** e **y**, deben corresponder a las muestras con las observaciones de las variables.

**Solo ocupar con menos de 5 variables**

### Prueba Chi-cuadrado Pearson

Sirve para inferir en las proporciones de dos variables categóricas y una de ellas es dicotómica (tiene solo 2 niveles), para ocupar el test hay que verificar

- Las observaciones deben ser independientes entre sí
- Deben haber al menos 5 observaciones en cada grupo

Existen 3 tipos de la misma prueba pero tiene diferencias conceptuales (no matemática)



## Prueba Chi-cuadrado de homogeneidad

Esta prueba resulta adecuada si queremos determinar si dos poblaciones (la variable dicotómica) presentan las mismas proporciones en los diferentes niveles de una variable categórica. Para determinar si las proporciones son similares y suficientes para proporcionar evidencia convincente, aplicaremos esta fórmula a cada casilla de la matriz, para luego armar otra matriz con los resultados a los cuales se le llamarán matriz de valores esperados

$$E_{(i,j)} = \frac{n_i n_j}{n}$$

donde:

- $n_i$ : es el total de observaciones en la fila  $i$
- $n_j$ : es el total de observaciones en la columna  $j$
- $n$ : tamaño de la muestra

Ahora para comparar las estadísticas podemos aplicar una fórmula para diferencia estandarizada la cual es la del modelo normal

$$Z = \frac{\text{estimador puntual} - \text{valor nulo}}{SE_{\text{estimador puntual}}}$$

En donde el error es la raíz cuadrada del valor nulo y el valor nulo es el valor de la posición  $(i, j)$  de la matriz de valor esperado y el estimador es el mismo punto  $(i, j)$  pero de la matriz original, luego consideraremos la suma al cuadrado de estos datos y así nace el estadístico  $\chi^2$

$$\chi^2 = \sum_{i=1}^m \sum_{j=1}^n Z_{ij}^2 = \sum_{i=1}^m \sum_{j=1}^n \frac{(\text{cantidad observada} - \text{cantidad esperada})^2}{\text{cantidad esperada}}$$

como  $\chi^2$  sigue una distribución chi-cuadrado, con  $\nu = (m - 1)(n - 1)$  grados de libertad

## Prueba chi-cuadrado de bondad de ajuste

Permite comprobar si una distribución de frecuencias observada se asemeja a una distribución de referencia, usualmente se emplea para comprobar si una muestra es representativa de la población

La forma de calcular las proporciones es la siguiente

$$P_i = \frac{n_i}{n}$$

donde:

- $n_i$ : es el valor de la columna  $i$ -ésima de la matriz
- $n$ : tamaño de la población

y luego la cantidad de personas esperadas es el resultado esperado multiplicado por el total de la muestra que se busca obtener

$$E_i = P_i n$$



- $n$ : tamaño de la muestra que se busca obtener

## Procedimiento POST-HOC

En la prueba de chi-cuadrado de proporción, la hipótesis nula se refiere a la igualdad de todas las proporciones, y la alternativa nos dice que no todas son iguales, pero no nos dice cuales, este tipo de pruebas son llamadas omnibus

Si una prueba de omnibus resulta significativa, existen metodos para identificar con mayor precision entre que tratamientos se dan estas diferencias, estos son llamados **procedimientos post-hoc** o **pruebas posteriori**, cabe destacar que **solo se hará un procedimiento post-hoc si la prueba omnibus rechaza la hipotesis nula en favor de la hipotesis alternativa**

En general las pruebas post-hoc realizan todas las comparaciones entre pares de tratameinto, si en una prueba se tienen  $k$  tratamientos, la cantidad de comparaciones ( $N$ ) es

$$N = \binom{k}{2} = \frac{k(k-1)}{2}$$

Para aplicar las pruebas hay que verificar el cumplimiento de las condiciones y mantener el mismo nivel de significancia de las pruebas post-hoc, la probabilidad de no cometer un error de tipo I en ninguna prueba es

$$P(\text{no error}) = (1 - \alpha)^n$$

donde  $n$  es la cantidad de pruebas realizadas, pero las pruebas de pares de tratmiento no son independientes.

la magnitud de la inflacion de la tasa de errores de tipo 1 depende de si las pruebas estan correlacionadas positivamente (diferencias significativas llevan a otras pruebas a tener diferencias significativas) o negativamente (diferencias significaitvas llevan a otras pruebas a no tener diferencias significativas)

Siempre es necesario ajustar los valores p obtenidos en las pruebas multiples para corregir la inflación de tasa de error 1, a esta probabilidad de cometer error de tipo 1 en al menos una de las pruebas se le llama 'tasa de error por familia de hipotesis' o FWER (family wise error rate en ingles)

El metodo más sensillo para corregir es el **metodo de bonferroni** que controla la FWER a un nivel  $\alpha$  rechazando solo la hipótesis nula cuyo valor p original es inferior a  $\alpha/m$ , alternativamemente se pueden obtener **valores p ajustados** ( $p_i^{\text{adj}}$ ), que se compara directamente con el nivel de significacion multiplicando el valor p obtenido con la cantidad de pruebas realizadas

$$p_i^{\text{adj}} = mp_i$$

esta correccion es poco usada por ser muy conservadora ya que mantiene los valores p abajo del valor de significancia, aumentando la probabilidad de cometer errores de tipo 2.

Otra alternativa es la **correccion de holm**, variante secuencial del metodo de bonferroni, pero este logra reducir su conservadurismo, el paso iterativo es, rechazar todas las  $H_0$  con valores p menores a  $\alpha/m$ , donde  $m$  es el numero total de comparaciones, esto deja  $m_1$   $H_0$  sin rechazar, luego rechazamos todas las que tengan valores p menores a  $\alpha/m_1$ , pero esta deja  $m_2$   $H_0$  sin rechazar, ... , y así, esto hasta que se rechacen todas las  $H_0$  o que un paso se complete sin rechazar ninguna.



El algoritmo se puede invertir para obtener valores  $p$  ajustados que se puedan comparar con el nivel de significancia, estos se obtienen:

$$p_i^{\text{adj}} = \max \left\{ (m - i + 1)p_{(i)}, p_{(i-1)}^{\text{adj}} \right\}, \text{ con } p_{(0)}^{\text{adj}} = 0$$

Así como estas pruebas son rigurozas para evitar errores de tipo 1, existen pruebas encargadas de evitar errores de tipo 2 en los casos donde se ocupan las pruebas anteriores en experimentos donde se realizan muchas pruebas y se hace más propenso el hecho de obtener falsos positivos, a estas pruebas se les llama tasa de falsos descubrimientos FDR (false discovery rate en inglés), una de las más conocidas es el **procedimiento de benjamin y hochberg** que presenta mayor potencia estadística que el resto de métodos basados en FWER, este método asume que los valores  $p$  de la familia son independientes o tienen una dependencia de regresión positiva, consiste en ordenar los  $m$  valores  $p$  de menor a mayor y rechazar todas las  $H_0$  cuyos valores  $p$  están en posiciones menores o iguales a  $(k)$ , tal que

$$k = \max_{i \in \{0, \dots, m\}} \left\{ p_{(i)} \leq \frac{i}{m} \alpha \right\}$$

si no existe la posición  $(k)$ , no se rechaza ninguna de las  $m$   $H_0$ , basándose en esta idea **benjamin y yekutieli** propusieron

$$k = \max_{i \in \{0, \dots, m\}} \left\{ p_{(i)} \leq \frac{i}{m} \frac{\alpha}{\sum_{i=1}^m \frac{1}{i}} \right\}$$

Con este cambio el procedimiento permite controlar la FDR al nivel  $(m_0/m)$ , donde  $m_0$  es el (desconocido) número de hipótesis nulas verdaderas, sin hacer suposición sobre la familia de valores  $p$ , pero este método es más conservador que el procedimiento original ya que los valores críticos se reducen por un factor  $\sum_{i=1}^m (1/i)$

En R ocuparemos la implementación provista por el paquete `stats`, en el cual la función `p.adjust(p, method)`, donde `p` es el vector de valores  $p$  y `method` es el argumento donde especificamos el tipo de método que queremos ocupar, donde tenemos las opciones `"bonferroni"`, `"holm"`, `BH` y `BY`.

## Procedimiento post-hoc para la prueba de chi-cuadrado

El procedimiento más simple consiste en efectuar múltiples pruebas chi-cuadrado entre cada par de tratamientos, si se verifican las condiciones o pruebas de Fisher, si las frecuencias esperadas se reducen mucho.

En R, existen pruebas ya implementadas para hacer esto, por ejemplo, `chisq.posthoc.test()` de paquete homónimo, `pairwiseNominalIndependence()` del paquete `rcompanion`, si se dan las condiciones también se pueden realizar múltiples pruebas de proporciones con la función `pairwise.prop.test(c, n, p.adjust.method)`

## Prueba de MCNEMAR

En esta prueba se considera el análisis de frecuencias apareadas, es decir cuando una misma característica con respuesta dicotómica se mide en dos ocasiones (o situaciones diferentes) para el mismo grupo de casos, la prueba busca determinar si se produce o no un cambio significativo en las proporciones observadas entre ambas mediciones



Los datos se pueden almacenar en una matriz de confusión como las que hemos visto en pruebas anteriores

		Variable 1		
		Presente	Ausente	Total
Variable 2	Presente	a	b	a+b
	Ausente	c	d	c+d
	Total	a+c	b+d	n

Las hipótesis que se manejan para esta prueba son:

- $H_0$ : no hay cambios significativos
- $H_a$ : si hay cambios significativos

dado que nos interesa medir cambios, solo nos sirven las celdas (c) y (b), así estas cuentan respectivamente éxitos y fracasos de una distribución binomial, cuando  $(b + c) > 10$  esta distribución se asemeja a una distribución normal con media  $\mu = (b + c)/2$  y desviación estándar  $\sigma = \sqrt{(b + c)/4}$ , a partir de la cual se puede obtener un estadístico  $Z$

Sin embargo en R se reporta el cuadrado de dicho estadístico, ignorando la condición de que sean 10 o más, el cual sigue una distribución chi-cuadrado con un grado de libertad

$$\chi^2 = \frac{(b - c)^2}{b + c}$$

dado que los datos siguen una distribución binomial que es discreta, pero se está usando como aproximación una chi-cuadrado entonces hay que aplicar un factor de corrección

$$\chi^2 = \frac{(|b - c| - 1)^2}{b + c}$$

si  $b + c < 25$  no es recomendable ocupar la prueba

Recordar que el valor p que obtenemos de la prueba es considerando una distribución chi-cuadrado, con un grado de libertad por lo que el valor p se calcula `pchisq(n, 1, lower.tail)`, en R esta prueba se puede utilizar como `mcnemar(x, Y=NULL, correct=FALSE)`, donde `y` se omite, `x` debe indicar la tabla de contingencia, de lo contrario, `x` e `y` deben ser las muestras con pares de observaciones de la variable dicotómica, `Correct` es para ocupar la versión con corrección o la sin corrección del método **Prueba Q de COCHRAN**

es una extensión de la prueba de mcneman, para cuando la variable de respuesta es dicotómica y la variable independiente tiene más de dos observaciones apareadas

Las hipótesis de la prueba son que si la proporción de éxitos es la misma (o no) en todas las mediciones, para realizar la prueba hay que verificar las siguientes condiciones:

- La variable de respuesta es dicotómica
- la variable independiente es categórica
- Las observaciones son independientes entre si



- el tamaño de la muestra es lo suficientemente grande,  $n_n n_t \geq 24$ , donde  $n_b$  es el numero de bloques que en que se organizan las observaciones y  $n_t$  es la cantidad de tratamientos estudiados

Un **bloque** es un caso que es sometido a las distintas mediciones, ej. una misma instancia que es resuelta por diferentes algoritmos

El estadistico de esta prueba se calcula así:

$$Q = \frac{\sum_{j=1}^{n_t} \left( x_{\bullet j} - \frac{N}{n_t} \right)^2}{\sum_{i=1}^{n_b} x_{i\bullet} (n_t - x_{i\bullet})} k(k-1)$$

donde:

- $n_b$ : cantidad de bloques
- $n_t$ : cantidad de tratamientos
- $x_{\bullet j}$ : total de exitos en la columna del j-esimo tratmiento
- $x_{i\bullet}$ : totla de exitos en la fila del i-esimo bloque
- $N$ : numero total de exitos

En R se utiliza la función `cochran.qtest(formula, data, alpha)` del paquete `RVAdMemoire`, donde

- `formula`: formula de la forma `respuesta tratamientos | bloques`
- `data`: matriz de datos en formato largo
- `alpha`: nivel de significancia

esta prueba considera una **hipotesis nula no parametrica omnibus**, la prueba post-hoc mas usada para este metodo es la de McNeman, En R `pairwiseMcNemar(formula, data, method)` nos permite aplicar esta prueba a cada par de tratamientos